

 NCBI

PubMed	Entrez	BLAST	OMIM	Taxonomy
Info				Structure

FASTA format description

- FAQs
- News
- References
- Credits

Education

- Program selection guide
- Tutorial
- URL API guide

Download

- Executables
- Databases
- Source code

Support

- Helpdesk
- Mailing list

A sequence in FASTA format begins with a single-line description, followed by lines of sequence data. The description line is distinguished from the sequence data by a greater-than (">") symbol in the first column. It is recommended that all lines of text be shorter than 80 characters in length. An example sequence in FASTA format is:

```
>gi|532319|pir|TVFV2E|TVFV2E envelope protein
ELRLRYCAPAGFAALLKCNDADYDGFKTNCNSNVVHCTNLMTNTVTTGLLNGSYSENRT
QIWQKHRTSNDSDLILNKHYNLTVTCKRPGNKTVLPVNAGLYFHQSQKYNLRLRQAWC
HFPSTMWKAWKEYKEEVNLPKERYRGNTDPKRIFEFQRQWGDEPTANLWFNCHGEFFYCK
MDWFELNYLNLTIVDADHNECKNTSGTKSGNKRAFGPCVQRTTYACHIRSVIWILETISKK
TYAPPREGHLECTSIVTGMTVYELNYIPKNRTNVTLSPQIESIWAELDRYKLVEITPIGF
APTEVRRYTGGHERQKRVPFVXXXXXXXXXXXXXVSQHLILAGILQQQNL
LAAVEBAQQQMLKLTIWGVK
```

Sequences are expected to be represented in the standard IUB/IUPAC amino acid and nucleic acid codes, with these exceptions: lower-case letters are accepted and are mapped into upper-case; a single hyphen or dash can be used to represent a gap of indeterminate length; and in amino acid sequences, U and * are acceptable letters (see below). Before submitting a request, any numerical digits in the query sequence should either be removed or replaced by appropriate letter codes (e.g., N for unknown nucleic acid residue or X for unknown amino acid residue).

The nucleic acid codes supported are:

A	-->	adenosine	M	-->	A	C	(amino)
C	-->	cytidine	S	-->	G	C	(strong)
G	-->	guanine	W	-->	A	T	(weak)
T	-->	thymidine	B	-->	G	T	C
U	-->	uridine	D	-->	G	A	T
R	-->	G A (purine)	H	-->	A	C	T
Y	-->	T C (pyrimidine)	V	-->	G	C	A
K	-->	G T (keto)	N	-->	A	G	C T (any)
			-				- gap of indeterminate length

For those programs that use amino acid query sequences (BLASTP and TBLASTN), the accepted amino acid codes are:

A	alanine	P	proline
B	aspartate or asparagine	Q	glutamine
C	cystine	R	arginine
D	aspartate	S	serine
E	glutamate	T	threonine
F	phenylalanine	U	selenocysteine
G	glycine	V	valine
H	histidine	W	tryptophan
I	isoleucine	Y	tyrosine
K	lysine	Z	glutamate or glutamine
L	leucine	X	any
M	methionine	*	translation stop
N	asparagine	-	gap of indeterminate length

Disclaimer
Privacy statement
Accessibility
Valid XHTML 1.0, CSS.